2014 International Conference on Future Internet of Things and Cloud

Exploiting Social Networks for the prediction of Social and Civil Unrest: A Cloud based Framework

Elhadj Benkhelifa; Elliott Rowe; Robert Kinmond; Oluwasegun A Adedugbe and Thomas Welsh
Faculty of Computing, Engineering and Sciences, Staffordshire University,
Stafford, ST18 0AD, UK
e-mail: e.benkhelifa@staffs.ac.uk

Abstract - **The current worldwide recession has led to a reduction in spending and a tightening of budget at all levels. Measures such as cuts in wages, lower pension pay-outs and rising unemployment seem to go hand-in-hand with politically motivated violence and social instability. In recent times, certain areas of Europe have been met with widespread protests, strikes and riots such as the ones in United Kingdom (UK), Spain, and Greece. Events over the last few years in the UK have demonstrated that people are willing to go to extreme lengths for their voice to be heard. Researchers in this area are still unclear about what leads to social instability and violent protests. How can these events be predicted? What tactics can be deployed by law enforcement agencies to manage these events? Social Networks such as Twitter and Facebook have been proven to be useful tools for demonstrators to organise themselves. Instead of limiting access to these services during any future disorders, filtered information fed from these media can be used by law enforcement agencies not only to prevent using them for criminal behaviour, but also to predict these events and develop tactics to manage future protests. This papers reviews the most cited research in this area and proposes a novel theoretical framework based on digital forensics principles combined with Cloud technology, followed by a sample practical implementation for illustration.**

Keywords — social networking; frameworks, cloud forensics; social network analysis, prediction, data exploration

I. INTRODUCTION

Riots and social unrest can happen for any number of reasons. They can be politically or economically motivated, or the result of a people's anger, frustration or upset [1]. Rioting can be observed throughout the course of history and in the modern age of social media, they are becoming more organised which makes them potentially more dangerous. In the last few years we have seen riots across the world in places as diverse as the UK, Egypt and Tunisia. These have been primarily politically motivated and have been a way for the people to let their governments know that they are unhappy.

Historically, riots have generally had a planned element to them. They are more organised now, partly as a result of the advent of social media [1]. The riots in the UK in 2011 are a prime example of the influence that social networks can have on the spread of rioting. The riots originated in the Tottenham area of London but soon spread to Birmingham, Manchester and many other areas of the UK.

The riots have been referred to as the 'BlackBerry Riots' [2] as they were organised primarily through the use of RIM's Blackberry Messenger service. Other social networks were also used during these riots. Twitter was used during and after the riots both to spread information about the riots and also in the aftermath to make remarks [2].

DataBlog [3], is a data gathering blog run by the UK newspaper, The Guardian. They have put together a chart during the time of the riots which shows the most popular tweets, the biggest trending topics and the Twitter accounts which had the most influence at the time of the riots. The data reveals the communication pattern which made the riots escalate and spread across the country very quickly. An area which law enforcement is keen to investigate is the possible use of social networks for prediction of the spread of unrest.

It has already been shown that social media can be used to give the public's general mood [4] or feeling at a certain time or about an event So it may be possible to use data from these media to identify the possible outbreak of a riot or civil unrest. Social networks and social media data have certain peculiar characteristics that other types of data do not possess. These characteristics mean that a framework to collect, analyse and utilise social media data to prevent civil unrests is vital and needs to be developed.

Some of the characteristics of social networks' data include:

- Data is updated in real-time and constantly changing.
- There are limits on the number of characters per message on some social media (e.g. twitter).
- The use of abbreviations, text speak and substitution of certain letters for numbers, which can affect the interpretation of a message while analysing it.

978-1-4799-4357-9/14 $31.00 © 2014 IEEE
DOI 10.1109/FiCloud.2014.98

- Website specific functions such as # or @ can provide extra information for data analysis.

These characteristics make analysing social media data and using it to predict events challenging. There are also other issues which may cause a social network investigation to be problematic, requiring adequate analysis through the use of a framework. Such issues include admissibility of evidence, perception of evidence, jurisdictional differences, forensic tools, lack of standards, volume of data, etc [5].

Cloud Forensics [6], a new discipline concerned with forensic investigation on digital content in the Cloud, share similar challenges and basic principles, as they require analysis upon datasets which have similar volatile and dynamic attributes. Therefore many of the problems found in Cloud Forensics are expected to be found in Social Network analysis. These challenges include loss of data control, no access to physical infrastructure, legal issues of multi-jurisdiction, multi-tenancy and multi-ownership, lack of tools for larger scale, distributed and virtualised systems, etc [6].

This papers reviews the most cited research in this area and proposes a novel theoretical framework based on digital forensics principles combined with Cloud technology, followed by a sample practical implementation for illustration.

II. RELATED WORKS

Several papers have analysed the impact social media plays in civil unrest within most communities. Ghannam [7] in his paper titled *'Social Media in the Arab World'* discusses the uprising and rioting in Tunisia and Egypt in 2011. He includes how the introduction of social media such as Twitter, Facebook, YouTube and for the Arab areas such as Aramram.com, 7iber.com, Ammannet.net, and AmmonNews.net have grown since the previous report was conducted in 2009.

The same issue is discussed by Dewel et al [8] in the paper '*The Impact of Social Media on Social Unrest in the Arab Spring*' in which he gave a more detailed overview of the issue as the specific details and quantitative results are used to make the author's point. In the paper '*Social Media as a tool for protest'* Papic et al [9] discussed the use of social media as a protest tool. Throughout they cite how important the use of online social networking has been, especially in the case of the 2011 Egyptian riots. The authors further referred to the revolutions in Iran and Moldova in 2009 as "The Twitter Revolution", thus identifying their links to social media.

According to Ghannam [7], between 2009 and 2010 there was a 12.5% rise in the number of active blog and bloggers in the Arab region. In this period, a team of 45 people from the Egyptian Interior Ministry were employed to monitor the 5 million active Facebook profiles in the region. These profiles belonged to amongst others, journalists, political leaders, human rights activists, social activists, entertainers and royalty who were engaging online in Arabic, English and French.

Safranek noted this in '*The Emerging Role of Social Media in Political and Regime Change'* [10] too by saying that the Middle East and North Africa region has one of the largest youth populations in the world, with 35-40% the population aged under 25. Ghannam [7] also discusses how the growth of social media has given the people free speech, and that digital communication technology has expanded the availability of tools for individuals to exercise their own freedom of expression, in the Arab region and other parts of the world as well.

Although Safranek [10] does not address this issue directly she notes how social media can break down boundaries and offer people the opportunity to be heard. Further, with the advent of the Internet and websites such as YouTube, Twitter and blogging websites, the need for credentials to be a source of information and opinion is lessened. This statement in the context of this article supports what Dewey et al [8] and Ghannam [7] both say about social media giving freedom of speech to all who use it.

Ghannam [7] then discusses how social media is beginning to replace traditional media in terms of the spread of news, noting that the number of Facebook users has already surpassed the number of newspapers sold in the region. Dewey et al [8] similarly noted that before the emergence of social media, stories of revolution and rioting were spread using traditional media, over which the authorities had some control. However the introduction of social networks has increased the speed at which news travels, and negated the control that the authorities have over the spread of information.

Ghannam [7] concludes that social media is having a direct impact on the nature of news delivery and engagement from the community and has given a voice to Arabs across the continent. Again both of these papers seem to agree that social media has been a positive factor in the spread of news in this region. Safranek [10] finishes by stating that the use of social media tools do not have one single outcome and trying to make close links with political action is often flawed. Whilst Papic et al [9] does not attribute full blame for protests on social networking, likening it to when cassette tapes of Ayatollah Khomeini were smuggled and used in the 1979 revolution in Iran. They do point out the vital role it plays and discuss the problem with the use of social networks.

A number of studies have been conducted on various Social Networks. For example, Theil et al [11] conducted a study looking at MySpace and also conducted a similar study on

Twitter. Some analyses are focused on specific events, such as the study conducted by Goh et al [12] which focused on people's response to the death of Michael Jackson, or the Tumasjan et al's study [13] on the political election in Germany. Other studies analysed broader social and economic trends, such as the relationship between Twitter mood and both stock market fluctuations, consumer confidence and political opinion [4]. This research showed that there was a link between social networks and politicalevents. Although not exact, these trends were able to give researchers an insight into the mood of a country at a given time. An additional factor noted was that performing these analyses online reduces the costs, efforts and time needed to administer large-scale public surveys and questionnaires.

A. Current Metthods and Techniques

There is almost no framework specifically relating to analyzing data on social media websites [14]. However there are a number of existing tools and techniques when combined together can create one. Whilst data mining can find connections in data, it does not always provide a context and does not take into consideration other variables. An activity or action that has a link to illicit or illegal activity does not mean that every person carrying out that activity or action will commit a crime [15]. Similar to data mining is text mining. Unlike data mining however where the intention is to find the "hidden" data, text mining is about finding the meaning in the known data [15]. Other methods and techniques include Big Data Analytics, NoSQL, Keyword Monitoring, Sentiment Analysis, Automated Semantic Role Labelling, etc [15].

B. Existing Framworks for Social Network Analysis

There are a number of generic digital forensic models which formed the basis for social network analysis frameworks. The most known and cited models are described below

Social Snapshot Framework - Because of the nature of social network data, normal frameworks may not be completely applicable to these types of investigations. Huber et al [16] have attempted to create a framework for investigations of this type. Huber's framework works on the principle of taking a social snapshot meaning that a snapshot of data will be taken at a given time. The diagram depicted in Figure 1 illustrates the model further.

Another framework worth mentioning is the one proposed by Zainudin et al [17] which is represented by Figure 2. The Zainudin et al model [18] seems to be in a constant loop meaning that even though one stage is performed before another, the stages can be repeated and referred back to.

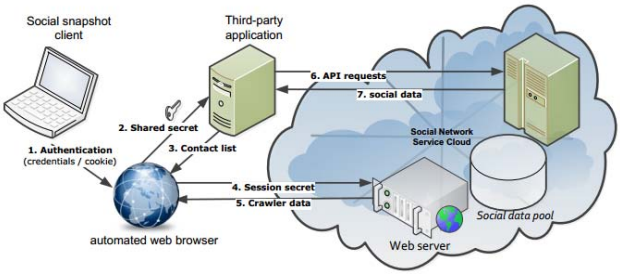

Fig. 1 Social Snapshot Framework [16]

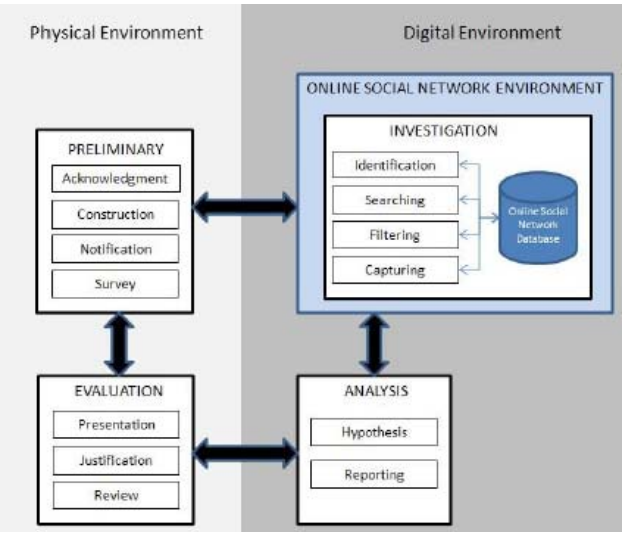

Fig. 2. Zainudin's Framework [17]

Other models also exist which though generic in nature, have been used within this domain. Some of them include

Abstract Digital Forensic Model - The Abstract Digital Forensic Model [19] applies natively to digital forensics. The actual stages of the process can be applied to both types of investigation. The stages in this model however are more detailed in terms of what each stage involves but there is still no real focus to what the stages are.

Integrated Digital Investigation Model - This model improves on the previous one (i.e. The Abstract Digital Forensic Model) and while it still uses the previous models as a basis it expands and builds on them, creating a more complete investigation framework [20]. The model is divided into different phases and there are specific contingencies for digital and traditional forensic investigations built into it. This model ensures that a more in-depth investigation is carried out.

Pollitt's Model - While Pollitt's model has a comprehensive structure, it does however suffer from being

generalised and not specific to the area in which it is used [21]. It does not give the user any details in regards to how the investigation should be carried out, and only gives a general overview to procedures. There are many issues to be considered when working with social network data, and as shown in the above section, many of the digital forensic frameworks that are currently in use do not take these issues into consideration. Therefore the proposed framework in the next section will attempt to fill in some of the gaps left by the current models.

Although there are already a number of different frameworks used for forensic investigations, many of them were not made with social network investigations in mind. Many of these frameworks are very generic, not appropriate for the dynamic nature of social network.

C. Current Analysis Tools

There are currently a variety of forensic data gathering and analysis tools, also used in social network analysis. In 2010, Raytheon developed a piece of software known as *RIOT (Rapid Information Overlay Technology)* which can be used as mining or even prediction software for social networks. The software takes data from social network sites and uses the data to provide information such as location at a specific time or date, commonly used words or phrases and the data can even be used to predict future events based on history [22].

Another similar tool to this is called *TouchGraph*. This software has multiple uses. It can be connected to Google, to find links between trends; it can also be connected to Amazon to see links between purchases and it can be connected to Facebook to show links between people on Facebook [23]. These types of software can be important in the tracking and monitoring of individuals and also the prediction of events. Furthermore, below is a selected and most common tools [24].

1. Analyst's Notebook - Part of the i2 Intelligence-Led Operations Platform, delivers rich assisted analysis and visualisation capabilities to support analysts to prevent crime and terrorism
2. Inflow - Software for social network analysis & organisational network analysis
3. Maltego - Open source intelligence and forensics application; allows identification of key relationships and networks
4. NodeXL - A free, open-source template for Microsoft Excel 2007 and 2010 that makes it easy to explore network graphs
5. GraphChi - Can run very large graph computations on just a single machine, by using a novel algorithm for processing the graph from disk (SSD or hard drive) [24].

III. PROPOSED FRAMEWORK

This paper presents a novel framework for the analysis and gathering of data from social networks and also the prediction of possible future incidents. The framework is built using the research carried out in the previous sections and will attempt to use some of the current frameworks, whilst attempting to add and improve them to create a more specific and focused framework for this type of investigation. Although the framework will follow a similar sequential pattern as the current models, it will also not be bound to the linear restrictions that the current models have. This framework will be more open and dynamic, similar to the data collected from social networks. The stages follow a sequence but they are created in such a way that they can be constantly looping while each subsequent stage takes place. This is because social data is so dynamic and it is vital to always have the most up to date information available. The proposed framework is divided into three major phases; the pre-investigation, investigation and the governance phases and is represented by the diagram below:

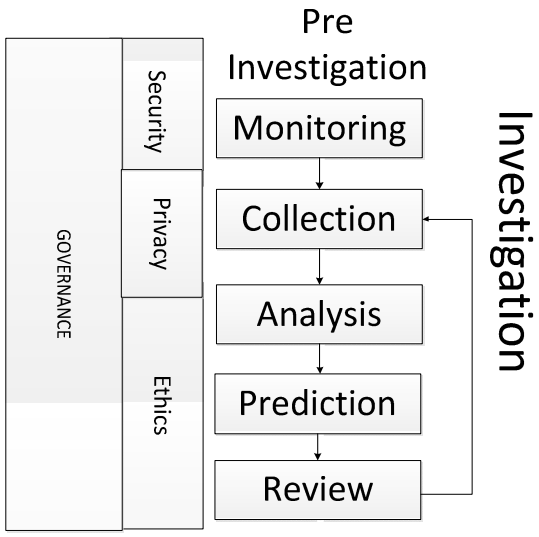

Fig. 3. Overall Proposed High Level Framework

A. Pre-Investigation Phase

Monitoring - This stage will be a constant repetitive phase throughout the entire investigation, meaning that the most up to date data will be available. The pre-investigation phase is where the initial situation monitoring will take place. During this phase, keyword monitoring will be set up to find any words or phrases which could be potentially

dangerous or of interest. The information will be gathered from the social network feeds of Facebook and Twitter, with particular attention being paid to feeds from areas which are known for outbreaks of violence. As well as the more obvious threat detection keyword monitoring, attention will be paid to the trending of place names and names of Government officials and any other important keywords. In addition to the keyword search, metadata such as time and source of the keyword will be recorded. When a keyword is flagged as potentially dangerous or of interest, the word will then be stored in a database for further investigation (collection phase). If the keyword is flagged, but has not appeared before, it will then be added to the keyword database. This way it will be constantly updated after any incident as shown in depicted in Figure. 4.

B. Investigation Phase

This phase is made up of sub-phases, which are collection, analysis, prediction and review. Collection - The evidence collection will be carried out using snapshot forensics. Due to the volatile nature of social media data, it is hard to gather every piece of data as it is constantly changing and updating. Using a snapshot will allow evidence from a particular time and data to be analysed. A social snapshot will be taken of the social network feeds, based on the information provided by the monitoring phase. After this the information will then be stored in a cloud database for analysis in the next stage. For this to be accurate, several snapshots must be taken over a set period of time. Calculating the best amount of time between each snapshot will be difficult however, as the data from Twitter and Facebook are dynamic and constantly updating, as seen in Figure 5.

Analysis - Once the data has been collected a full analysis will be carried out on it. The data will be analysed using processes such as Sentiment Analysis and Automated Semantic Role Labelling [25]. This will give the investigators more information into the possible meaning behind the messages and provide a context. Another tool used here will be Aggarwal's framework [26] for the collection and analysis of data from traditional sources. The data will be analysed and stored in a database in the cloud. The database will be run using Hadoop due to the size of the data, as illustrated in Figure 6.

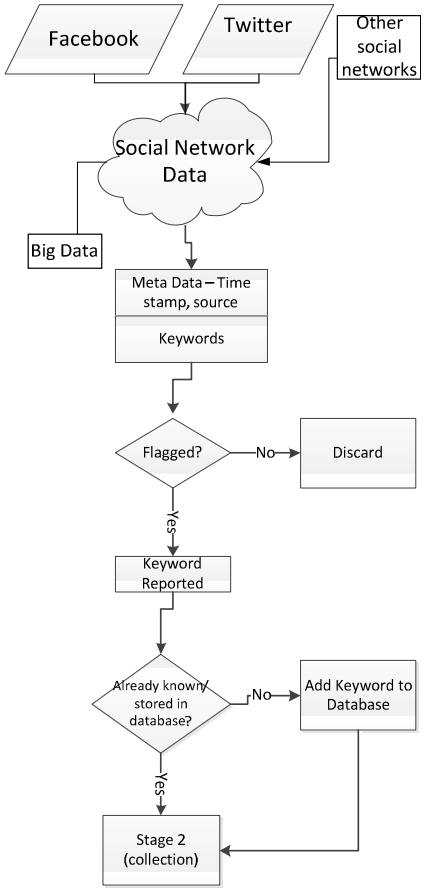

Fig. 4. Proposed Monitoring Framework

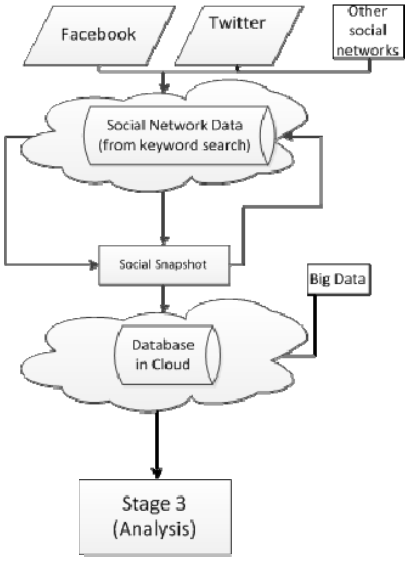

Fig. 5. Proposed Collection Framework

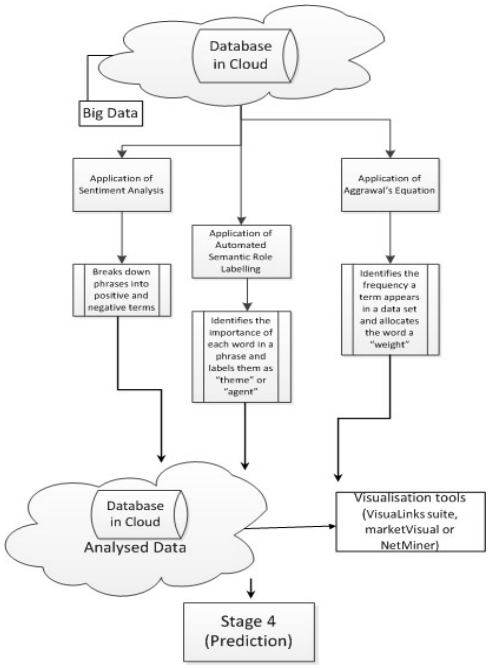

Fig. 6. Analysis Framework

Prediction - This is the stage in which techniques are put into use to predict events. Keyword monitoring from social network feeds will be used again to gather the information, but this information will be compared to a database which contains the activity of social networks during times of peace and of unrest.

Raytheon's RIOT software [22] can be used to predict what the outcome of certain data sets will be based on a previous data set. Therefore this can be used to predict what trends are likely to result in the outcome of unrest or rioting. Data Mining techniques such as forecasting can be applied during this stage of the investigation. Specific data mining frameworks such as the Information Fusion-based Indication and Warning Assessment, Recognition System (IIWARS) can be applied at this stage as seen in Figure. 7.

Review - After the data has been compared to previous data sets in the prediction stage, it many need further investigation. This stage will then loop back to the collection phase allowing the investigator to collect more information on this specific data set.

C. Governance Phase

Within the framework, when working with data pulled from social networking sites, there are several important factors which must be taken into consideration. These include:

Privacy - Within an investigation of this nature, it is likely that excess information will be obtained. Any information that is not found to be relevant to the investigation must be treated with privacy and confidentiality, with the details being kept secure. The EC-Council guidelines state [27].

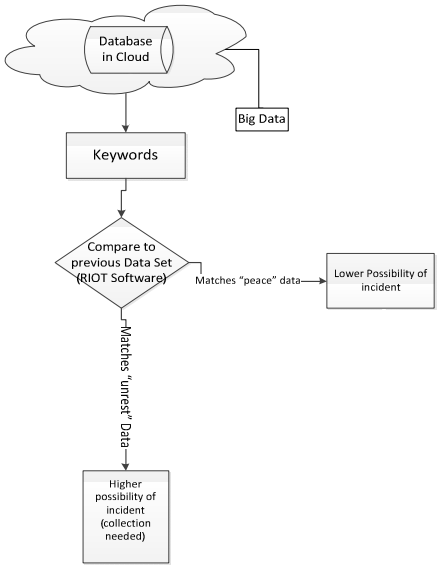

Fig. 7. Proposed Prediction Framework

Ethics - Similarly to privacy, an investigation of this nature can be filled with many ethical concerns. It is up to the investigator to treat all information found, whether related to the investigation or not, in an ethical way. There are many different versions of the ethical guidelines that digital forensic investigators abide by. One example is the code of ethics laid down by the EC-Council in reference to ethical hackers.

Abundant Information - Not all the information gathered from the social networks will be relevant to a case. In some cases however, relevant information will be dismissed as it will not be picked up by the keyword searches. In some cases, such as Twitter, characters like "#" and "@" have unique properties when used on the site. In these cases, special attention will need to be paid by investigators so as not to miss important data.

Unstructured Data - Similarly, the keyword searches may struggle to decipher certain text due to the words being abbreviated, shortened or even certain letters being replaced by numbers, for example "ate" and "8". This can be solved in the same way by updating a database with each new case.

Short Length - Like the abundant information problem, websites such as Twitter limit the amount of characters in a single tweet to 140. This means that in some cases, letters will be removed like the use of "u" instead of "you". This can be resolved in the same way by having an active

checking system and slowly building a database which will have these characteristics.

Time Sensitivity - Due to the constantly changing flow of data from social networks it will not always be possible to have exact up to date information. Therefore social snap-shotting will be used. This method will allow data from certain time intervals to be captured.

Visualization - In order to make it easier to display the information, once collected, software such as VisuaLinks Suite, MarketVisual or NetMiner can be used to display the results in the form of a graph. This will enable the investigator to be able to analyse the data in an easier to read and more manageable format. The use of link analysis software will also mean the investigator can see links between keywords and how they correlate.

Big Data Storage - Due to the size of the data collected, as previously discussed big data management techniques will be in place in this framework. The data collected will be stored in an online cloud using Hadoop and a NoSQL database.

IV. Proof-of-Concept Implementation

A proof-of-concept implementation was developed and executed in order to test and prove the framework. The implementation involved a dynamic and scalable cloud-based system which monitored twitter for terms related to civil unrest. Once these terms were noticed, information regarding places and locations were extracted and then their rate and sentiment were monitored further to provide a dataset for predictive analytics. Further technical details about this implementation can be found in a recently published paper by the authors [18]

A. Arhictecture

The system used a multi-agent cloud-based architecture, which dynamically scaled according to load; this was suited to the attributes of social media networks. The system composed of two types of agents, social media monitoring agents and sentiment analysis agents (Figure .8).

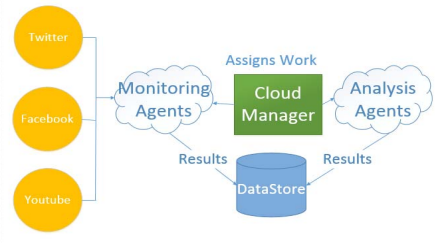

Fig.8 Platform architecture.

B. Initial Test

An initial test was conducted on the term "Woolwich" due to ongoing incidents within the UK. The system extracted a number of place names and dates which may relate to civil unrest events and then built a graph relating these terms (Figure 9). These terms were then monitored as a time series showing posts and their sentiments in order to provide data to support predictive analytics of civil unrest, an example of a sentiment stream is given in Figure 10.)

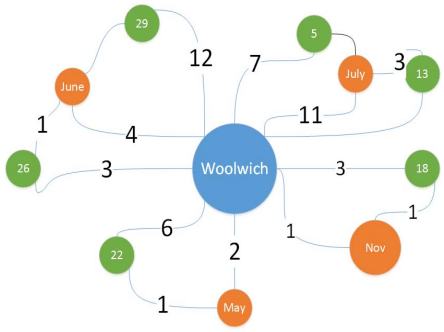

Fig. 9 – A graph structure showing extracted potential dates for civil unrest related to the "woolwich" incident.

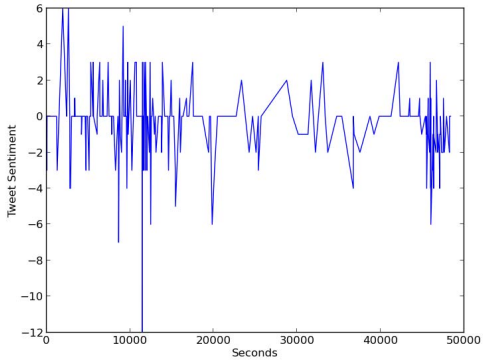

Figure. 10 – An example time series analysis, showing rate and sentiments of a given term

V. CONCLUSION

As online social networks become intertwined with modern day life, their effect on the physical world becomes greater every day as such, this paper examined their role in civil unrest. Due to the similarity in monitoring online social media and digital cloud forensics, a framework was developed which spawned as a combination of these. This framework may be applied in order to analyses such networks and produce datasets in order to potentially predict new incidents of civil unrest. To validate this framework a proof-of-concept implementation was given,

which monitored twitter for signs of civil unrest in order to determine potential locations and dates.

VI. REFERENCES

[1] H. Ismailov (2011). "England Riots: What can Literature teach us?". BBC [online]. Available: http://www.bbc.co.uk/blogs/legacy/worldservice/writerinresidence/2011/08/england_riots_what_can_literat.html. [Accessed 29 April, 2013]

[2] I. Mackenzie (2011, Aug. 8). *"Is Technology to blame for the London riots?"*. BBC News [Online]. Available: http://www.bbc.co.uk/news/technology-14442203. [Accessed: March. 27, 2012].

[3] L. Evans and F. Vis. (2011). "200 most influential Twitter users during the riots: are you on the list?". 08 Aug. 2011. *DataBlog*. Available http://www.guardian.co.uk/news/datablog/2011/dec/08/riot-Twitter-top-200 [Accessed: 30 Apr 2013]

[4] J. Bollen, et al., Modelling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena. 2010.

[5] J. Son. Social Network Forensics: Digital Forensics for Social Media- Evidence Extraction tools capabilities. Lap Lambert Academic Publishing. 2012.

[6] K. Ruan, "Cloud Forensics: an Overview" Cloud Futures, Microsoft Research. 2012.

[7] J. Ghannam. "Social Media in the Arab World: Leading up to the Uprisings of 2011" Centre for International Media Assistance, 2011.

[8] T. K. Dewey. "The Impact of Social Media on Social Unrest in the Arab Spring". Final Report prepared for: Defence Intelligence Agency 2012.

[9] M. Papic et al. "Social Media as a tool for protest". Security Weekly. 2011.

[10] R. Safranek. "The Emerging Role of Social Media in Political and Regime Change". ProQuest Discovery Guides. 2012.

[11] K. Thiel et al. "Creating Usable Customer Intelligence from Social Media Data: Network Analytics meets Text Mining". 2012.

[12] D. H. Goh et al. "An analysis of tweets in response to the death of Michael Jackson". *Emerald Group Publishing Limited*. 2011.

[13] A. Tumasjan et al. "Predicting Elections with Twitter". *Fourth International AAAI Conference on Weblogs and Social Media.* 2011.

[14] A. Java et al. "Why We Twitter: Understanding Microblogging Usage and Communities". *NEC Laboratories America.* 2010.

[15] W. J. Seifert. "Data Mining and Homeland Security: An Overview" CRS Report for Congress. 2007.

[16] M, Huber et al. "Social Snapshots: digital forensics for online social networks". SBA Research, 2011.

[17] N, Zainudin et al. "A digital forensic investigation model for online social networking". School of Computing and Mathematical Sciences Liverpool John Moores University. 2008.

[18] E.Benkhelifa and T. Welsh (2014). A Cloud Enabled Social Media Monitoring Platform for Events Detection and Prediction. *Internet Technology and Secured Transactions (ICITST), 2013 8th International Conference for. Page(s):142 – 148.*

[19] M. Reith. "An Examination of Digital Forensic Models". *International Journal of Digital Evidence*. vol 1, issue 3. 2002.

[20] B. Carrier and E. H. Spafford. "Getting Physical with the Investigative Process". *International Journal of Digital Evidence*. Volume 2, Issue 2. 2003.

[21] M, Pollitt. "An Ad Hoc review of digital forensic models". *Proceedings of the Second International Workshop on Systematic Approaches to Digital Forensic Engineering*. 2007.

[22] Raytheon (2011). "*Raytheon Software Tracks*". The Guardian [online]. Available: http://www.guardian.co.uk/world/video/2013/feb/10/raytheon-software-tracks-online-video [Accessed: 30 Apr 2013]

[23] TouchGraph (2012). [online] Available: http://www.touchgraph.com, 2012. [Accessed: 28 Apr 2013].

[24] KDNuggets (2012). [online] Available : http://www.kdnuggets.com/software/social-network-analysis.html. [Accessed: 28 Apr 2013].

[25] S. W. Yih and K. Toutanova. *Automatic Semantic Role Labelling.* Microsoft Research. 2007.

[26] A, Aggarwal et al. "Systematic digital forensic investigation model". *International Journal of Computer Science and Security (IJCSS)*, Volume (5): Issue (1), 2011.

[27] EC-Council. (2010). *Code of Ethics* [online] Available : https://www.eccouncil.org/ciso/becoming-cciso/code-of-ethics. [Accessed: 28 April 2013].

[28] R, Miller. (2010). "Facebook Now Has 30,000 Servers". [online] Available: http://www.datacenterknowledge.com/archives/2009/10/13/Facebook-now-has-30000-servers/, [Accessed: 29 April 2013].

[29] Facebook Server Count [online] available: http://www.datacenterknowledge.com/archives/2010/06/28/Facebook-server-count-60000-or more/, Feb 2013.